Uncertainty Quantification of Density and Stratification Estimates with Implications for Predicting Ocean Dynamics

A. MANDERSON,[a] M. D. RAYSON,[b] E. CRIPPS,[a] M. GIROLAMI,[c] J. P. GOSLING,[d] M. HODKIEWICZ,[e] G. N. IVEY,[b] AND N. L. JONES[b]

[a] *Department of Mathematics and Statistics, University of Western Australia, Perth, Western Australia, Australia*
[b] *Oceans Graduate School, University of Western Australia, Perth, Western Australia, Australia*
[c] *Department of Mathematics, Imperial College London, and Alan Turing Institute, British Library, London, United Kingdom,*
[d] *School of Mathematics, University of Leeds, Leeds, United Kingdom,*
[e] *Faculty of Engineering and Mathematical Sciences, University of Western Australia, Perth, Western Australia, Australia*

(Manuscript received 7 November 2018, in final form 6 March 2019)

ABSTRACT

We present a statistical method for reconstructing continuous background density profiles that embeds incomplete measurements and a physically intuitive density stratification model within a Bayesian hierarchal framework. A double hyperbolic tangent function is used as a parametric density stratification model that captures various pycnocline structures in the upper ocean and offers insight into several density profile characteristics (e.g., pycnocline depth). The posterior distribution is used to quantify uncertainty and is estimated using recent advances in Markov chain Monte Carlo sampling. Temporally evolving posterior distributions of density profile characteristics, isopycnal heights, and nonlinear ocean process models for internal gravity waves are presented as examples of how uncertainty propagates through models dependent on the density stratification. The results show 0.95 posterior interval widths that ranged from 2.5% to 4% of the expected values for the linear internal wave phase speed and 15%–40% for the nonlinear internal wave steepening parameter. The data, collected over a year from a through-the-column mooring, and code, implemented in the software package Stan, accompany the article.

1. Introduction

The horizontal and vertical distribution of the buoyancy frequency

$$N(z) = \left(-\frac{g}{\rho_0} \frac{d\rho}{dz} \right)^{1/2} \tag{1}$$

and hence the background density $\rho(z)$ is one of the most important dynamical characteristics of the ocean (Phillips 1977). The background density (or buoyancy frequency) is a key input variable in many ocean process models including planetary wave propagation (Gill 1982), internal gravity wave energy flux calculations (Nash et al. 2005; Lee et al. 2018), nonlinear internal wave models (Lamb and Yan 1996), strain-induced turbulent dissipation parameterizations (Polzin et al. 2014), length-scale-based vertical mixing parameterizations (Arthur et al. 2017), and internal wave available potential energy calculations (Kang and Fringer 2010). Such models require a precise estimate of the background density profile, uncontaminated by internal waves and turbulent fluctuations, and, importantly, a profile that spans the entire water column from the free surface to the seabed.

In low and midlatitudes, the vertical density (temperature) structure usually consists of a surface mixed layer, a sharp gradient region (the pycnocline or thermocline), and then a weaker gradient region below where the change in density increases (temperature decreases) at an exponentially lower rate (Sprintall and Cronin 2010). An exception to this general picture are so-called fossil layers that form double pycnoclines, usually found in subtropical regions (Sprintall and Roemmich 1999). The pycnocline region is the most dynamically and ecologically significant aspect of the ocean stratification because the strong gradients suppress turbulent transport of tracers, yet there is still considerable debate around how to objectively quantify its main properties, such as the depth and strength (Fiedler 2010). Fiedler (2010) provides an overview of several historical strategies for determining upper-ocean thermal

Corresponding author: M D. Rayson, matt.rayson@uwa.edu.au

DOI: 10.1175/JTECH-D-18-0200.1

structure, including fitting a series of discontinuous linear segments.

Continuous analytical functions are an effective means of capturing the vertical density, or temperature, structure because they allow properties such as pycnocline depth (depth of maximum gradient) and strength (maximum gradient) to be calculated objectively (e.g., Chu et al. 1997; González-Pola et al. 2007; Zhou et al. 2017). González-Pola et al. (2007) propose the product of a Gaussian and an exponential to represent the temperature structure below the surface mixed layer, whereas Zhou et al. (2017) use a series of Gaussian curves. González-Pola et al. (2007) cites the physical basis for their function: it approximately resembles a complementary error function, a solution to a one-dimensional heat diffusion equation under appropriate boundary conditions. An exponential profile is often used to represent the density structure in the deeper ocean, although it does not capture the pycnocline variability (e.g., Llewellyn Smith and Young 2003). More recently, Rayson et al. (2019) use a double hyperbolic tangent (DHT) function to reconstruct the full vertical structure of $\rho(z)$ in the upper ocean observed from mooring data. The DHT function is flexible enough to model two pycnoclines as it approximately resembles two error functions, and its parameter values provide direct inference about density profile characteristics. In these studies, profile parameters were estimated via a minimization algorithm [Zhou et al. (2017) use least squares fitting, González-Pola et al. (2007) use an evolutionary algorithm, and Rayson et al. (2019) use robust least squares] and thus only providing a single estimate of each density profile.

The above works produce valid databased estimates of density profiles but offer no mathematically coherent mechanisms to quantify the degree of belief or uncertainty that accompanies any such synthesis of hypothetical model and data. Probability is perhaps the most acknowledged calculus for uncertainty quantification and scientific inference (Singpurwalla 2006; Hennig et al. 2015), but concepts, such as confidence intervals and standard errors, of traditional frequentist statistics do not supply a probabilistic description of parameter uncertainty (Jaynes and Kempthorne 1976). This work demonstrates how to formally account for uncertainty by uniting the DHT of Rayson et al. (2019) with a probabilistic framework that is suited to the structure of data analyzed herein. Before introducing how this is accomplished, we discuss why this is important in the context of density profile estimation and, importantly, the implied consequences for ocean dynamical process models (i.e., nonlinear functions of density) that are analyzed in more detail later.

First, our knowledge of a true, full-water-column, continuous ocean background profile density will always be uncertain. Estimates are typically derived through an equation of state using temperature–conductivity measurements informed either from moored instruments at a finite number of depths, often without measurements near the surface or bottom boundaries, or from repeated vertical profiles with continuously sampling instruments that seldom resolve down to the seabed (Nash et al. 2005). Natural variability also contributes to noisy fluctuations at various time scales because of seasonal processes (e.g., mean ocean currents, monsoons), weekly processes (e.g., mesoscale eddies), and hourly processes (e.g., internal gravity waves). For internal gravity wave propagation, we are interested in $\rho(z)$ at time scales greater than the inertial period ($\gtrapprox$1 day) and the choice of smoothing splines, temporal filters, and finite-difference schemes to produce estimates of $\rho(z)$, and hence $N(z)$, greatly vary between studies (e.g., Chelton et al. 1998; Nash et al. 2005; King et al. 2012). These are density profile–specific issues, to which may be added more general concepts such as measurement error, model misspecification, and sparse data. For an excellent review on uncertainty sources and oceanography, see Wikle et al. (2013, 2019).

Second, it is important to recognize that the uncertainty associated with $\rho(z)$ has implications for models of ocean dynamical processes that depend upon $\rho(z)$. To clarify, consider the following two ocean dynamic examples considered throughout this article. First, a simple baroclinic ocean process model describing motion in a flat-bottom ocean in a motionless background state with the quasigeostrophic approximation is

$$\frac{d^2\phi(z)}{dz^2} + \frac{N(z)^2}{c^2}\phi(z) = 0, \tag{2}$$

subject to boundary conditions $\phi(0) = \phi(-H) = 0$, where $-H$ is the total water depth (Gill 1982). Since computation of $N(z)$ relies on $\rho(z)$, so too does the vertical structure of each normal mode $\phi(z)$ and the phase speed c. Any uncertainty associated with the initial value of $\rho(z)$ will lead to uncertainty in the function ϕ.

As a second example, the Korteweg–de Vries (KdV) equation,

$$\frac{\partial A}{\partial t} + c\frac{\partial A}{\partial x} + \alpha A\frac{\partial A}{\partial x} + \beta\frac{\partial^3 A}{\partial x^3} = 0, \tag{3}$$

where

$$\alpha = \frac{3c\int_{-H}^{0}(d\phi/dz)^3\,dz}{2\int_{-H}^{0}(d\phi/dz)^2\,dz}\,(\mathrm{s}^{-1}), \tag{4}$$

is often used to model the evolution of nonlinear internal waves in shelf seas as an alternative to the computationally expensive Navier–Stokes equations (e.g., Holloway et al. 1999). If $\alpha < 1\,\mathrm{s}^{-1}$, this corresponds to waves of elevation, and for $\alpha > 1\,\mathrm{s}^{-1}$, waves of elevation. In lieu of solving Eq. (3) directly, Rayson et al. (2019) demonstrate that the initial growth of nonlinear internal waves is suitably estimated by a steepening length scale derived from a balance between the time-dependent and nonlinear terms in Eq. (3):

$$L_s = \frac{c^2}{\alpha a_0 \omega}\,(\mathrm{m}), \tag{5}$$

where a_0 and ω are the initial wave amplitude and frequency, respectively. Equation (5) suggests that nonlinear waves may occur if L_s is less than the distance from an internal wave generation site (e.g., a continental shelf slope or an island chain). But L_s is proportional to c and inversely proportional to α, both of which are nonlinear functions of $\rho(z)$, and thus uncertainty in specifying $\rho(z)$ creates uncertainty in L_s.

This article quantifies uncertainty using probability theory and performs statistical inference via Bayes's law. Bayes's law is a framework that, in receipt of new information such as data, coherently updates a *prior* distribution over unknown quantities (e.g., summarizing prior scientific knowledge) to their *posterior* distribution (Jeffreys 1998). The posterior distribution then supplies direct probabilistic inference and uncertainty quantification, conditional on the data. When the process of interest is complex, Bayesian hierarchical models (BHMs) can be used to decompose the problem into a series of conditionally independent components, the product of which yields the model's joint distribution. Compartmentalizing the model both eases the burden of specifying dependencies in high-dimensional distributions and is a flexible way to meld multiple sources of information. BHMs are well suited to multivariate data that are repeatedly measured (e.g., spatiotemporal data), where local processes within a measurement benefit from the sharing of global information supplied by repetition (Gelman et al. 2013; Betancourt and Girolami 2015).

Previous applications of BHMs to oceanography can be found in a number of papers: Berliner et al. (2000) use BHMs to produce long-lead sea surface temperature predictions, Furrer et al. (2007) combine observations and future climate projections within a BHM to predict atmosphere–ocean circulation processes, Milliff et al. (2011) develop a regional ocean forecast system for surface vector wind fields, and Aldrin et al. (2012) assess climate sensitivity using hemispheric temperatures and global ocean heat content data. By embedding prior knowledge of ocean dynamics within a probabilistic framework, such work resides at the interface of statistical and oceanographic research with analytically unavailable posterior distributions, often due to high-dimensional and strongly dependent parameter spaces, that are estimated using carefully constructed Markov chain Monte Carlo (MCMC) sampling algorithms (Metropolis et al. 1953; Hastings 1970). For a more comprehensive review of BHMs applied to oceanographic processes, their challenges, and estimation strategies, see Wikle et al. (2013).

The dataset we consider here contains multivariate density profile measurements over depth in the upper ocean. Repeatedly collected at regular time intervals over a year, each profile measurement is regressed on depth at the first level of the BHM, using DHT functions as the expected values. While each DHT function has its own parameterization, the second level specifies that these parameters a priori share a common probability generating mechanism. This permits not only the borrowing of global (or annual) information to reduce local (or time specific) parameter uncertainty but also incorporates the evolution of the parameters, and hence $\rho(z)$, over the year. Where domain knowledge is available, the third level specifies priors derived from expert elicitation (O'Hagan et al. 2006); otherwise, priors that are weakly informative are used. Finally, the posterior distribution of $\rho(z)$ that is implied by the BHM can be propagated through models of ocean dynamical processes to quantify their uncertainty.

Although the BHM is easily interpretable and models the evolution of profile density measurements well, the statistical nonlinearity of the DHT and the complex spatiotemporal dependencies contained in its joint distribution induce a posterior distribution sufficiently complicated to appeal to more sophisticated and efficient MCMC algorithms. We use Hamiltonian Monte Carlo (HMC), an algorithm that exploits geometric knowledge about the posterior surface (Betancourt 2017). The model is implemented in the open-source software Stan (Carpenter et al. 2017). Diagnostic plots to verify the MCMC algorithm, as produced by Stan, are presented in the article. Data, code, and additional documentation are publicly available online (https://github.com/hhau/density-profile-examples). The R package used to fit the density profile model is available separately (https://github.com/hhau/ddcurves2). Quantification

of uncertainty is presented via the following estimated quantities: full posterior distributions of DHT parameters at selected time points and posterior intervals for their corresponding density profiles and the annual evolution of posterior intervals of density profile characteristics, isopycnal heights, and the nonlinear functions for the ocean dynamical processes $N(z)^2$, $\phi(z)$, c, α, and L_s.

The paper is structured as follows. Section 2 gives a brief introduction to Bayesian inference and MCMC algorithms, which is intended for the nonspecialist statistician. Section 3 describes the DHT function in the context of marginal seas, and section 4 describes the shelf mooring dataset used as an example in this paper. Section 5 contains details of the BHM, while section 6 presents the results. Section 7 concludes the article.

2. Bayesian inference and estimation via MCMC

a. The Bayesian paradigm

Statistical modeling begins with an assumed data generating process that gives rise to a probabilistic object called the *likelihood function* $p(\mathbf{y}|\boldsymbol{\theta})$: the probability (density or mass) function of the observed data in $\mathbf{y}$, conditioned on the model parameters contained in $\boldsymbol{\theta}$. In frequentist statistics, where probability is defined in a limiting frequency sense, $\boldsymbol{\theta}$ is considered unknown but fixed. The uncertainty associated with any estimator of $\boldsymbol{\theta}$ (e.g., maximum likelihood) is quantified by the estimator's probability distribution over notional repeated samples, but does not provide direct probabilistic inference for a given observed sample. Under the Bayesian paradigm, where probability is a quantitative representation of belief or uncertainty, our initial knowledge about $\boldsymbol{\theta}$ is encoded in a *prior* distribution $p(\boldsymbol{\theta})$. In light of the observed data, invoking Bayes's theorem updates $p(\boldsymbol{\theta})$ to the *posterior* distribution,

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{6}$$

The current degree of belief is now summarized in $p(\boldsymbol{\theta}|\mathbf{y})$ (Bernardo and Smith 1994). Uncertainty associated with the data is described in $p(\mathbf{y}|\boldsymbol{\theta})$, and $p(\boldsymbol{\theta})$ allows the incorporation of additional sources of information, from previous studies to domain-specific knowledge.

Some researchers find it difficult to articulate conclusions drawn from frequentist methods, for example, the interpretation of a confidence interval that hinges on hypothetical repeated samples (Jaynes and Kempthorne 1976). In contrast, Bayesian inference is appealing because Eq. (6) supplies direct probabilistic statements about model parameter uncertainty, conditioned on the particular dataset observed. For example, estimates of $p(\boldsymbol{\theta}|\mathbf{y})$ and 0.95 credible intervals (a, b), which correspond to a posterior probability $p(a < \boldsymbol{\theta} < b|\mathbf{y}) = 0.95$, are used in section 6 to quantify uncertainty.

b. Posterior estimation

The advent of MCMC sampling methods and the increasing availability of computational power has enabled Bayesian statistics to tackle complex statistical problems. Prior to MCMC, the difficulty with these endeavors was that the normalizing constant in Eq. (6), $p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\theta})\, d\boldsymbol{\theta}$, was very likely intractable. MCMC avoids this issue as follows: construct a Markov chain of the model parameters $\boldsymbol{\theta}^{[0]}$, $\boldsymbol{\theta}^{[1]}$, $\boldsymbol{\theta}^{[2]}$, ... proposing a value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$, for the tth realization from a suitable proposal density, $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t-1]})$. Accepting this proposal with probability

$$\min\left\{1, \frac{p(\boldsymbol{\theta}^*|\mathbf{y})q(\boldsymbol{\theta}^{[t-1]}|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{[t-1]}|\mathbf{y})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{[t-1]})}\right\} \tag{7}$$

yields a transition kernel that guarantees the Markov chain will converge to a stationary distribution that coincides with $p(\boldsymbol{\theta}|\mathbf{y})$. Importantly, the ratio in Eq. (7) obviates the need to compute the normalizing constant. By simulating such a Markov chain until convergence, the subsequent realizations may be used to estimate $p(\boldsymbol{\theta}|\mathbf{y})$. Furthermore, these MCMC samples can also be used to estimate the posterior distributions of functions of model parameters, such as the ocean dynamic quantities presented in section 6.

In theory, the choice of $q(\cdot)$ in Eq. (7) is extremely flexible, but in practice, $q(\cdot)$ has strong implications for the efficacy of MCMC algorithms for two main reasons. First, MCMC theory ensures asymptotic convergence but not convergence in a finite number of simulated realizations. Second, as opposed to ordinary Monte Carlo analysis, these realizations are not independent but may be heavily autocorrelated depending on the choice of $q(\cdot)$. Quantitative diagnostics for the first issue are fragile, and it is common to visually validate that multiple chains, initialized from different states, converge to the same stationary distribution. The idea is then to discard the "warm up" realizations prior to convergence, and use the remainder for posterior estimation. The autocorrelations can be checked with standard autocorrelation function (ACF) plots and ought not display dependencies at unacceptably high lags. To establish an approximate independent sample size that corresponds to the MCMC output, the *effective sample size* can be computed as

$$\mathrm{ESS} = \frac{n}{1 + \sum_{k=1}^{\infty} r(k)}, \tag{8}$$

where n is the number of MCMC samples used for estimation and $r(k)$ is their autocorrelation at lag k. These diagnostics for our analysis are presented in section 6. For an introduction and more details about MCMC and its validation diagnostics, we direct the interested reader to Robert and Casella (2013) and Brooks et al. (2011).

The complexity of modern day applied Bayesian analyses requires input from both domain specialists and statisticians. One outcome of these collaborations is statistical research into the construction of efficient MCMC proposal densities due to parameter spaces that are high-dimensional and/or exhibit strong posterior dependencies [see Girolami and Calderhead (2011) and discussions]. In these situations, standard MCMC algorithms, such as random-walk Metropolis–Hastings (Hastings 1970) or Gibbs sampling (Gelfand and Smith 1990), can result in excessive warm-up periods and unacceptably strong autocorrelations. HMC, a recent variant of MCMC, is specifically designed to deal with such issues. HMC proposal densities harness first-order gradient information about the posterior surface to identify regions of sampling exploration and ensure high acceptance rates. While the theoretical foundations of HMC, and how it circumvents the above issues, are beyond the scope of the present article, we direct the interested reader to the textbook chapter of Neal (2011), a conceptual introduction by Betancourt (2017), and a short overview aimed at applied scientists in Monnahan et al. (2017).

c. *Overview of Bayesian hierarchical models*

BHMs have been greatly transformed in general by MCMC sampling (Gelman and Hill 2007; Gelman et al. 2013) and more recently by its variant HMC (Betancourt and Girolami 2015; Monnahan et al. 2017). BHMs are a model-building strategy in which data and unknown parameters are organized into a series of conditionally independent models, thereby easing the strain of building large multivariate models with complex covariance structures. Suppose that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in Eq. (6), then a BHM may decompose the problem by

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) &= p(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \\ &= \underbrace{p(\mathbf{y}|\boldsymbol{\theta}_1)}_{\text{Stage 1}}\,\underbrace{p(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)}_{\text{Stage 2}}\,\underbrace{p(\boldsymbol{\theta}_2)}_{\text{Stage 3}}, \end{aligned} \tag{9}$$

where Eq. (9) indicates that, conditional on $\boldsymbol{\theta}_1$, $\mathbf{y}$ is independent of $\boldsymbol{\theta}_2$. Equation (9) is a succinct summary of the full BHM proposed in this paper: stage 1 describes the data, conditional on the parameters that govern the ocean pycnocline model parameters, stage 2 provides the mechanism to allow the pycnocline model parameters to vary temporally, and stage 3 specifies priors on the remaining parameters contained in $\boldsymbol{\theta}_1$.

Fortunately, the appeal of BHMs has led to the development of many software packages (also known as probabilistic programming languages) to facilitate MCMC implementation for the scientific community. While users must understand the basics of Bayesian modeling and MCMC output, these software packages provide straightforward model specification, data synthesis, and output retrieval. This article uses Stan (Carpenter et al. 2017) and its interface to R (R Core Team 2018), RStan (Stan Development Team 2018). The HMC algorithm inside Stan (Hoffman and Gelman 2014) is also adaptive, so hand tuning of the proposal density by the practitioner is minimal, and it readily provides the MCMC diagnostics suggested above. However, we note that Stan does have one drawback that more traditional MCMC algorithms and accompanying software do not: the inability to estimate models with discrete (or noncontinuous) parameters because of its need to compute gradients (Gelman et al. 2015).

3. A parametric ocean pycnocline model

Ocean water density $\tilde{\rho}(z, t)$ is derived from temperature, salinity, and pressure variables, measured at discrete depths from either a vertical mooring or a profiling instrument, through an equation of state. A filtering operation is typically applied to the time series to extract the background density from the measured (raw) data via $\rho(z) = \langle\tilde{\rho}(z, t)\rangle$, where $\langle\cdot\rangle$ indicates a low-pass-filter operation. It is the background density that is the dynamically significant quantity of interest for most analyses.

A parametric model for the density profile is needed to reduce the dimensionality from a finite number of discrete points to a few parameters that can provide inference. Such a model should have the following properties. First, it must appropriately fit observed data in the presence of either one or two distinct pycnoclines, the most dynamically significant aspect of the ocean stratification. Second, the function must allow for sensible extrapolation both above the shallowest observation to the ocean surface and beyond the deepest observation to the ocean floor. Such behavior is not an immediate attribute of parametric forms such as polynomials nor nonparametric methods such as splines or Gaussian processes. Third, the function should be monotonic to maintain gravitational stability.

We use a DHT model similar to Liu and Benney (1981) and Stastna and Lamb (2002) that suitably captures the pycnocline structure in marginal seas and

the upper ocean in low and midlatitudes (depths less than 500 m). Liu and Benney (1981) present their model in terms of buoyancy frequency,

$$N(z)^2 = N_1^2 \mathrm{sech}^2\left(\frac{z+\beta_2}{\beta_3}\right) + N_2^2 \mathrm{sech}^2\left(\frac{z+\beta_4}{\beta_5}\right). \quad (10)$$

Via integration, a DHT function for this type of background density is

$$\rho(z) = \beta_0 - \beta_1\left[\tanh\left(\frac{z+\beta_2}{\beta_3}\right) + \tanh\left(\frac{z+\beta_4}{\beta_5}\right)\right]. \quad (11)$$

The buoyancy frequency derived from Eq. (11) is similar in form to Liu and Benney (1981), where $N_1^2 = -g/\rho_0(\beta_1/\beta_3)$ and $N_2^2 = -g/\rho_0(\beta_1/\beta_5)$. Equation (11) also satisfies the requirements mentioned previously and has the added benefit that each parameter has a physical interpretation as illustrated in Fig. 1. Figure 1 depicts how the first parameter β_0 (kg m^{-3}) is approximately the mean density over the profile, while β_1 (kg m^{-3}) is a scale for the density difference over the water column. The additive factors β_2 and β_4 (m) are the middepths of the upper and lower pycnoclines, respectively, and the denominators β_3 and β_5 (m) are the pycnocline widths. Properties such as the pycnocline depth, defined as the depth of maximum gradient, and the strength, defined as the maximum gradient, can now be computed exactly from Eq. (10). Finally, note that Eq. (11) may also be used to represent the vertical temperature structure by changing the sign in front of β_1 and noting that β_0 and β_1 would have units of degrees Celsius.

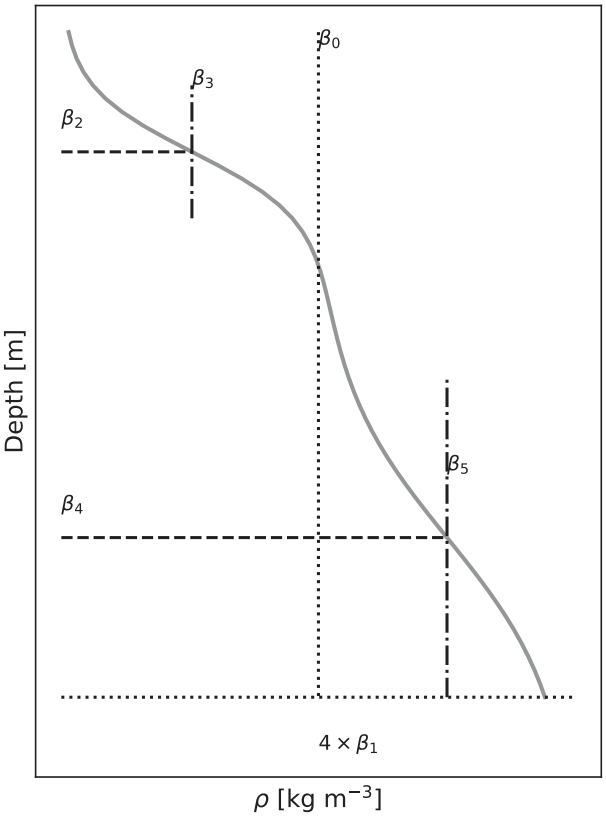

FIG. 1. Schematic of the parametric double hyperbolic tangent (DHT) density profile model [Eq. (11)].

4. Mooring data

We demonstrate the suitability of embedding the parametric form in Eq. (11) within a BHM using temperature data collected from a through-the-water-column mooring deployed in 250-m-deep shelf region on the Australian North West Shelf (13.7589°S, 123.3462°E,) for 12 months (April 2016–May 2018). The mooring was serviced in early November 2016 to download data and remove biofouling, resulting in two 6-month data blocks separated by a 2-day gap. The mooring consisted of 11 thermistors (Starmon mini temperature loggers) spaced at 20–30-m-depth intervals and sampling every 60 s. The uppermost instrument was nominally located 10 m below the free surface, although knockdown due to drag caused this depth to vary. A pressure sensor on the upper float measured knockdown, allowing us to map temperature onto a moving vertical coordinate. Temperature data were converted to density by assuming constant salinity (34.6 psu) and a nonlinear equation of state (Feistel 2008). Raw or measured density $\tilde{\rho}(z)$ was converted to background density $\rho(z)$ by first linearly interpolating the instrument depths onto a constant depth array, then using a backward-in-time low-pass filter with a cutoff time scale of 34 h (Rayson et al. 2019).

Background density for the 12-month record was linearly interpolated from the discrete observation depths to produce a time series of isopycnals heights (Fig. 2). The density field in the region evolved over an annual time scale with smaller, weekly time-scale fluctuations. A cool mixed layer down to approximately 70-m depth occurred during the *austral* winter (August 2016) with a sharp pycnocline between 70 and 200 m. Lighter, more buoyant water appeared in the upper 50 m during the summer months because of warming and the pycnocline also broadened. A secondary pycnocline occurred intermittently throughout the 12-month period. However, this view (Fig. 2) of the data does not contain any information about the uncertainty of the isopycnal heights or about general features like the pycnocline width. In section 6, we show how our BHM produces not only point estimates of the isopycnal heights, as reported in Fig. 2, but also their credible intervals.

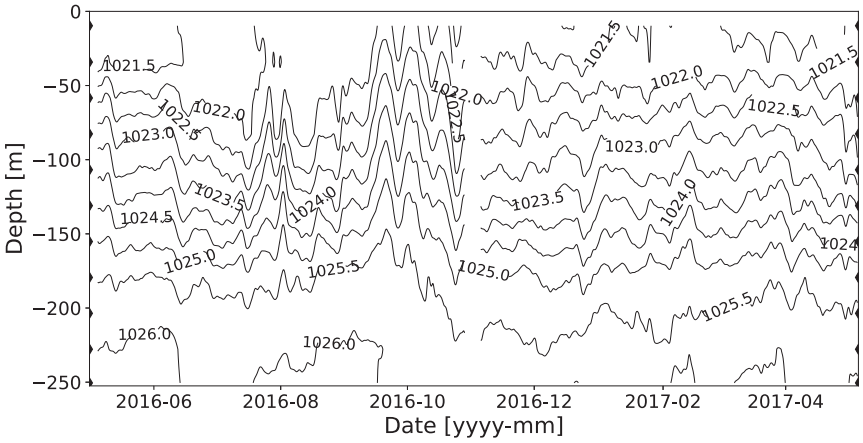

FIG. 2. Background density temporal evolution derived from the mooring data. Nominal measurement heights are indicated by black diamonds. Contours (isopycnals) are spaced at 0.5 kg m^{-3} intervals and computed via linear interpolation.

5. The Bayesian hierarchical model

As in section 4, we have a mooring at a fixed latitude/longitude with a one-dimensional spatial component over depth and repeated measures over time. The first level of the hierarchy describes each time-specific observed density vector (over depth) as a DHT function with normal errors. The second stage allows the parameters of the DHT to vary from profile to profile according to a multivariate normal distribution. The third stage completes the BHM by specifying prior distributions for the remaining parameters and also utilizes scientific knowledge about the characteristics of vertical density profiles.

a. Data

Let $y_{d,t}$ for $t = 1, 2, \ldots, T$ and $d = 1, 2, \ldots, D$ be the observed seawater density at time t and depth d. For the data in this article, observations correspond to $T = 1473$ time periods and $D = 11$ discrete depths. Write $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{D,T})'$, $\mathbf{z} = (z_1, z_2, \ldots, z_D)'$, then at time t, our BHM specifies

$$\begin{aligned}\mathbf{y}_t &= \beta_{0,t} - \beta_{1,t}\left[\tanh\left(\frac{\boldsymbol{z}+\beta_{2,t}}{\beta_{3,t}}\right) + \tanh\left(\frac{\boldsymbol{z}+\beta_{4,t}}{\beta_{5,t}}\right)\right] + \boldsymbol{\varepsilon}_t \\ &= \rho(\mathbf{z}, \boldsymbol{\beta}_t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, \sigma_y^2 \mathbf{I}_D),\end{aligned} \tag{12}$$

where all arithmetic operations involving vector quantities occur in an element-wise manner, $\mathbf{I}_D$ is the $D \times D$ identity matrix, and the errors contained in the $D \times 1$ vector $\boldsymbol{\varepsilon}_t$, are independent normal random variables with zero mean and variance σ_y^2. By allowing the regression parameters $\boldsymbol{\beta}_t = (\beta_{0,t}, \beta_{1,t}, \ldots, \beta_{5,t})'$ to vary across time, Eq. (12) captures the local temporal characteristics of the density profiles. Writing $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_T')'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_T')'$, the likelihood function is

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma_y^2) = \prod_{t=1}^{T} (2\pi\sigma^2)^{-D/2} \exp\left\{-\frac{1}{2\sigma^2}[\mathbf{y}_t - \rho(\mathbf{z}, \boldsymbol{\beta}_t)]'[\mathbf{y}_t - \rho(\mathbf{z}, \boldsymbol{\beta}_t)]\right\}. \tag{13}$$

b. Density model parameters

Although the values of $\boldsymbol{\beta}_t$ in Eq. (12) vary temporally, we assume they are generated by a common probability distribution. This is also an opportunity to utilize prior knowledge. We incorporate this into the distribution of $\boldsymbol{\beta}_t$ as

$$\boldsymbol{\beta}_t \sim N_+(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}) \quad \text{for} \quad t = 1, 2, \ldots, T, \tag{14}$$

where $\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\mu_{\beta_0}, \mu_{\beta_1} \ldots, \mu_{\beta_5})'$, $\Sigma_{\boldsymbol{\beta}} = \text{diag}(\sigma^2)$, and $\boldsymbol{\sigma}^2 = (\sigma_{\beta_0}^2, \sigma_{\beta_1}^2, \ldots, \sigma_{\beta_5}^2)$. The notation N_+ indicates a normal distribution constrained to be positive and hence ensures the density profiles at each time point are monotonic. Equation (14) implies the existence of some common density profile parameterized by $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and that the density profile at a specific time t is a deviation from this mean. This level of the BHM structure allow us to share information across time points, permits global inference about $\boldsymbol{\beta}_t$, and results in a more precise estimate of σ_y (Gelman et al. 2013). The BHM model described until this point is depicted as a directed acyclic graph in Fig. 3.

c. Final level priors

To complete the Bayesian specification of the model, we require prior distributions on all remaining model

parameters. The following priors for the regression coefficient means are the result of an expert elicitation process (Astfalck et al. 2018) about the possible behavior of vertical density profiles. Given the physical interpretation of these parameters, as discussed in section 3 and shown in Fig. 1, our priors that appropriately encode the expert knowledge are

$$\mu_{\beta_0} \sim N_+(1025, 10^2), \quad \mu_{\beta_1} \sim N_+(5, 2^2),$$
$$\mu_{\beta_3} \sim N_+(80, 15^2), \quad \mu_{\beta_5} \sim N_+(80, 15^2),$$
$$\mu_{\beta_2} \sim N_+(75, 15^2), \quad \mu_{\beta_4} \sim N_+(150, 15^2).$$

The choice of prior distributions for the mean parameter incorporates our knowledge about reasonable values of certain parameters. For example, scientific knowledge suggests it is unlikely that the mean value of the background ocean density over the whole water column falls outside the range of 1005 to 1045 kg m^3. Thus, our prior for μ_{β_0}, the typical value of the density profile intercept term, places 0.95 of the prior (not posterior) probability mass in this interval. The prior distributions for the remaining parameters contained in $\boldsymbol{\mu_\beta}$ have been elicited in a similar manner.

On the other hand, prior specification of standard deviation parameters is more difficult to elicit. Furthermore, if the number of repetitions within the data is small or the standard deviations approach close to zero, uninformative priors, despite their intention, can exert a significant influence on the posterior (Gelman 2006). Our dataset contains substantial repetitions ($T = 1473$), and we specify weakly informative priors similar in spirit to those recommended by Gelman et al. (2013). In particular, we set $\sqrt{\sigma^2_{\beta_i}} = \sigma_{\beta_i} \sim N_+(0, 15^2)$ for $i = 0, 1, \ldots, 5$ and $\sqrt{\sigma^2_y} = \sigma_y \sim N_+(0, 0.25^2)$. While these priors specify more likely (positive) values toward zero, they taper off as the standard deviation increases, allowing the posterior to be dominated by the likelihood rather than the prior, if necessary.

6. Results

We first present some diagnostics of the MCMC output produced by Stan. Since the BHM has 8851 unknown parameters, we restrict this discussion only to $\boldsymbol{\mu}_{\beta_i}$, $i = 0, 1, \ldots, 5$, and note that the remaining parameters display similar output. One chain of length 5500 samples takes 1.6 h on a standard workstation hardware (4-core Intel i5 at 3.3 GHz with 8 GB of RAM). The first 2000 samples are discarded as the warm-up period. Three chains, each initialized

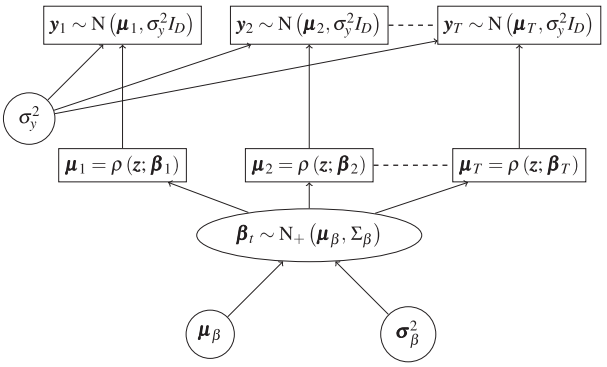

FIG. 3. Directed acyclic graph of the hierarchical model. Circular nodes are (hyper)parameters of the model, for which we specify prior distributions. The elliptical node is the structural node that enforces the common prior for each $\boldsymbol{\beta}_t$. Rectangular nodes are either derived quantities from the model parameters and DHT model or observations.

from different states, are presented in Fig. 4 for the post-warm-up period and have converged to the same stationary distribution. The ACF plots in Fig. 5 do not exhibit significant autocorrelations at unacceptably large lags. Across all model parameters, the minimum effective sample size as calculated by Eq. (8) is 3337 samples. As a result, we consider the MCMC algorithm to be satisfactory for estimating the posterior distribution.

In describing the BHM results and the effect of uncertainty quantification on ocean dynamics, we first demonstrate the appropriateness of the BHM described in section 5 by comparing $p[\rho(z, \boldsymbol{\beta}_t)|\mathbf{y}]$ with density profile observations at three time periods chosen to exhibit different stratification conditions. For the same time periods, $p[N^2_t(z)|\mathbf{y}]$ and $p[\phi_t(z)|\mathbf{y}]$ are discussed. Next, the annual evolution (i.e., for $t = 1, 2, \ldots, T$) of $p(c_t|\mathbf{y})$, $p(\alpha_t|\mathbf{y})$, and $p(L_{s,t}|\mathbf{y})$ illustrate the uncertainty in the linear baroclinic phase speed, nonlinear internal wave parameter, and internal wave steepening length scale given in Eqs. (2), (4), and (5), respectively. Finally, $p(z^*_t|\mathbf{y})$, where z^*_t is an isopycnal height that satisfies $\rho^* - \rho(z^*_t, \boldsymbol{\beta}_t) = 0$ for a given density ρ^*, provides a novel way to illustrate isopycnal contours. All posterior distributions are summarized by expected values, $\int xp(x)\,dx$, and 0.95 credible intervals (a, b) such that $p(a < X < b) = \int_a^b p(x)\,dx = 0.95$. In what follows, the statements "expected value" and "credible interval" refer to these summaries of the relevant posterior distribution.

Posterior distributions of the more complicated functions are estimated using the MCMC output. As an example, $p[N^2_t(z)|\mathbf{y}]$ can be estimated by evaluating

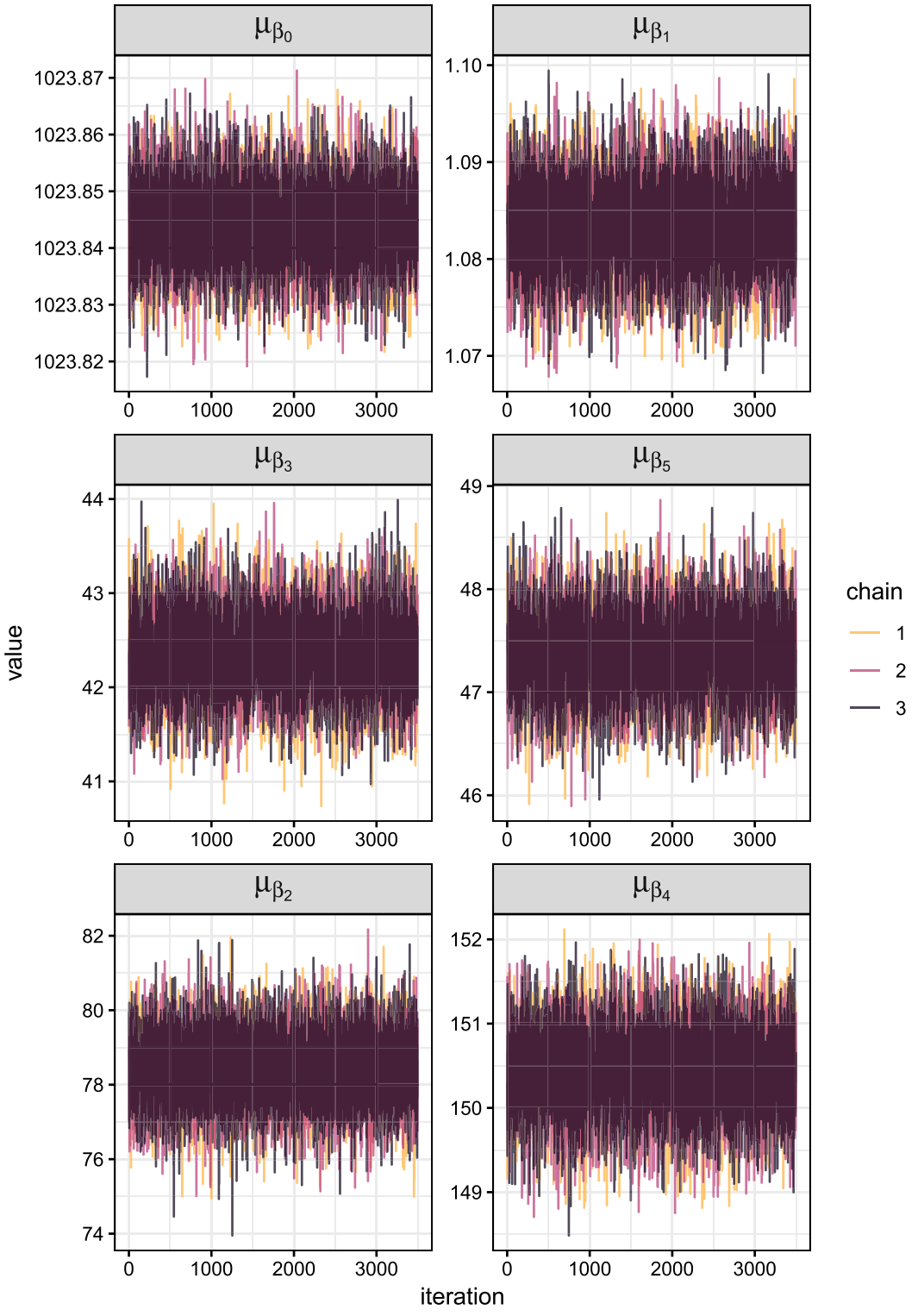

FIG. 4. Trace plots of μ_{β_i}, $i = 0, 1, \ldots, 5$, for three MCMC chains simulated by Stan, each of length 3500 and post-warm-up period. Individual chains were initialized with different values.

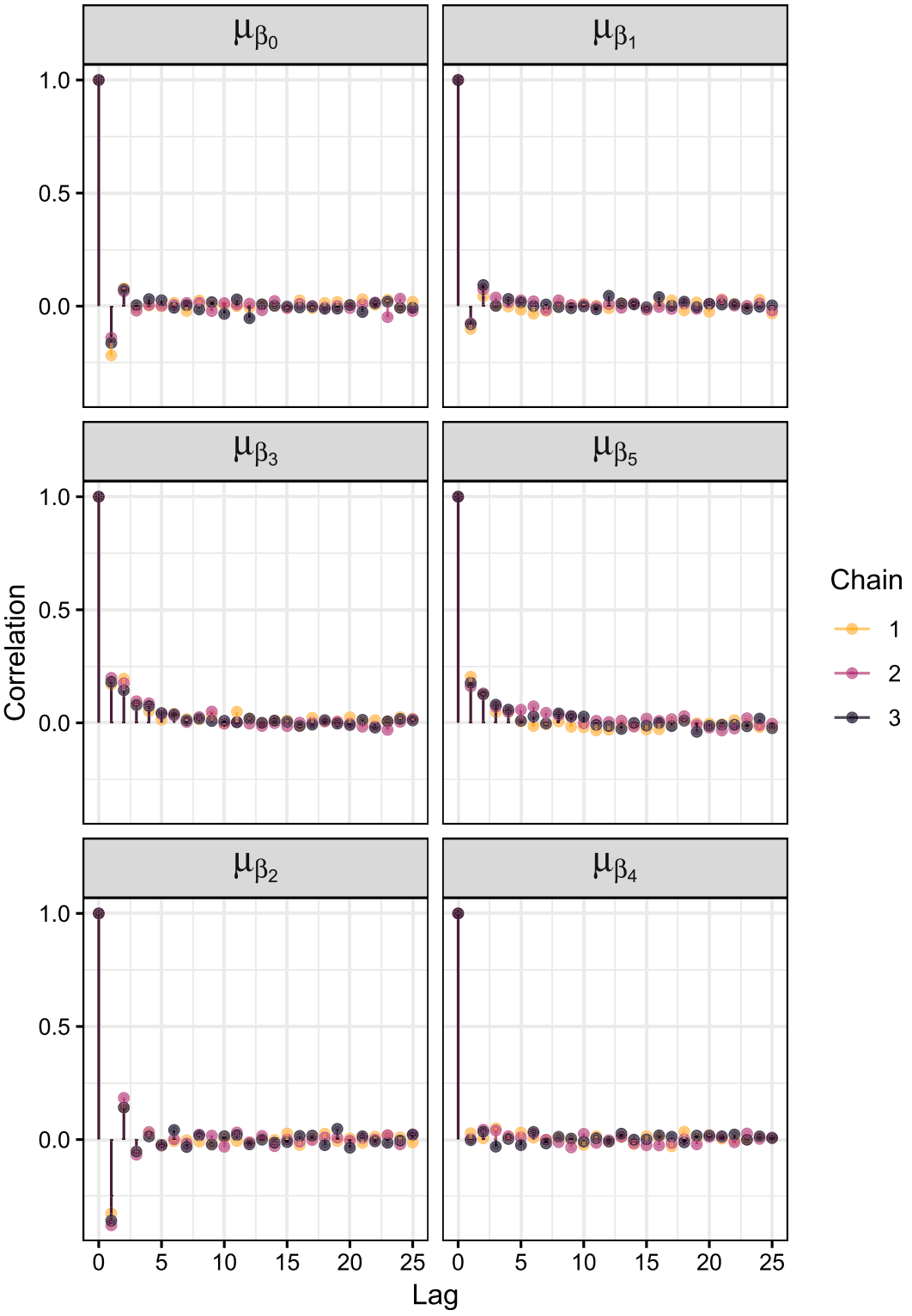

FIG. 5. ACF plots of μ_{β_i}, $i = 0, 1, \ldots, 5$, for three MCMC chains simulated by Stan, each of length 3500 and post-warm-up period.

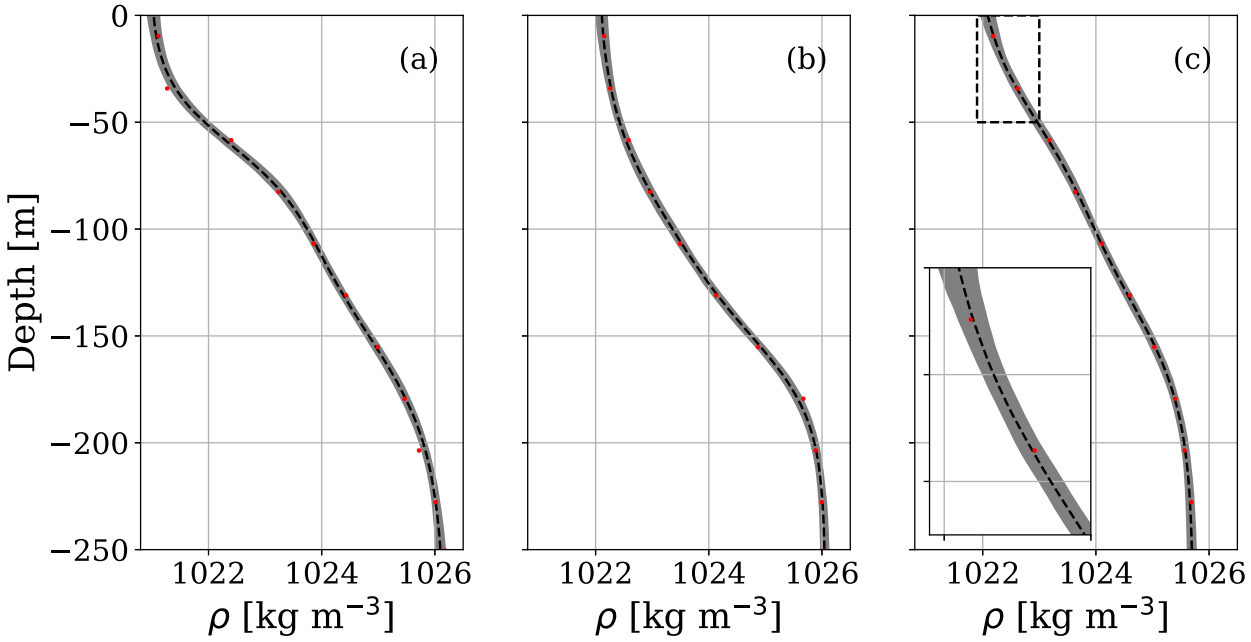

FIG. 6. The posterior mean (line) and 0.95 credible intervals (gray shading) for the double hyperbolic tangent density profile model. Red dots indicate the background density profile measurements. (a) Plot for 7 May 2016 ($t = 27$) when the density profile was characterized by was a double pycnocline. (b) Plot for 26 Aug 2016 ($t = 469$) when there was a single pycnocline. (c) Plot for 9 Oct 2016 ($t = 647$) when there was a broader single pycnocline.

$$N_t^2(z) = \frac{g\beta_{1,t}}{\rho_0}\left[\beta_{3,t}^{-1}\operatorname{sech}^2\left(z + \frac{\beta_{2,t}}{\beta_{3,t}}\right) + \beta_{5,t}^{-1}\operatorname{sech}^2\left(z + \frac{\beta_{4,t}}{\beta_{5,t}}\right)\right] \quad (15)$$

for MCMC samples of $\boldsymbol{\beta}_t$ and treating these as realizations from $p[N_t^2(z)|\mathbf{y}]$. Similarly for $p(c_t|\mathbf{y})$, $p(\alpha_t|\mathbf{y})$, $p(L_{s,t}|\mathbf{y})$, and $p(z_t^*|\mathbf{y})$.

a. Background density profile and characteristics

Three example time periods, shown in Fig. 6, demonstrate the performance of the BHM for density profiles that exhibited diverse stratification conditions: a double pycnocline on 7 May 2016 ($t = 27$), a single pycnocline on 26 August 2016 ($t = 469$), and a broader single pycnocline on 9 October 2016 ($t = 647$). Figure 6 shows that the expected values fit the data well, providing sensible monotonic interpolation and extrapolation both at the sea bed and surface. The credible intervals for the density profiles in Fig. 6 were greater at the sea surface where the profile was extrapolated upward. Overall, the credible interval widths ranged from 0.1 to 0.25 kg m^{-3}.

Figure 7 displays the posterior distributions of the profile characteristics for the three example time periods in Fig. 6. Recall that the characteristics provided by the DHT function are the mean density β_0, density difference across the water column β_1, and the two pycnocline depths β_2 and β_4 and widths β_3 and β_5. The expected value for $\beta_{1,t}$ was largest during the double pycnocline period (Fig. 6a; 6 May 2016; $t = 27$) with a value of 1.3 kg m^{-3}. This date corresponded to expected values of $\beta_{2,t}$ and $\beta_{3,t}$ of 59.8 and 30.7 m, respectively. In contrast, single pycnocline time periods showed larger expected values for $\beta_{3,t}$ of 53 m at $t = 469$ and 44.4 m at $t = 647$ (Figs. 6b,c), indicating wider upper pycnoclines. Furthermore, both single pycnocline time periods reported an expected value of the density difference $\beta_{1,t}$ of 1.0 kg m^{-3}. The relatively small density difference and relatively wide upper pycnocline were the key characteristics that resulted in the single pycnocline structure as opposed to the distinct double pycnocline seen at $t = 27$ (Fig. 6a).

Figure 7 demonstrates how each of the DHT parameters behave for the three contrasting density profiles and posterior distributions of the parameters for all time points are presented in Fig. 8. The expected values of the $\beta_{0,t}$s had a maximum of 1024.1 kg m^{-3} on 9 September 2016 ($t = 563$) with a credible interval of (1023.9, 1024.3). The minimum expected value of the $\beta_{0,t}$s was 1023.2 kg m^{-3} on 19 November 2016 ($t = 801$) with a credible interval of (1023.1, 1023.3).

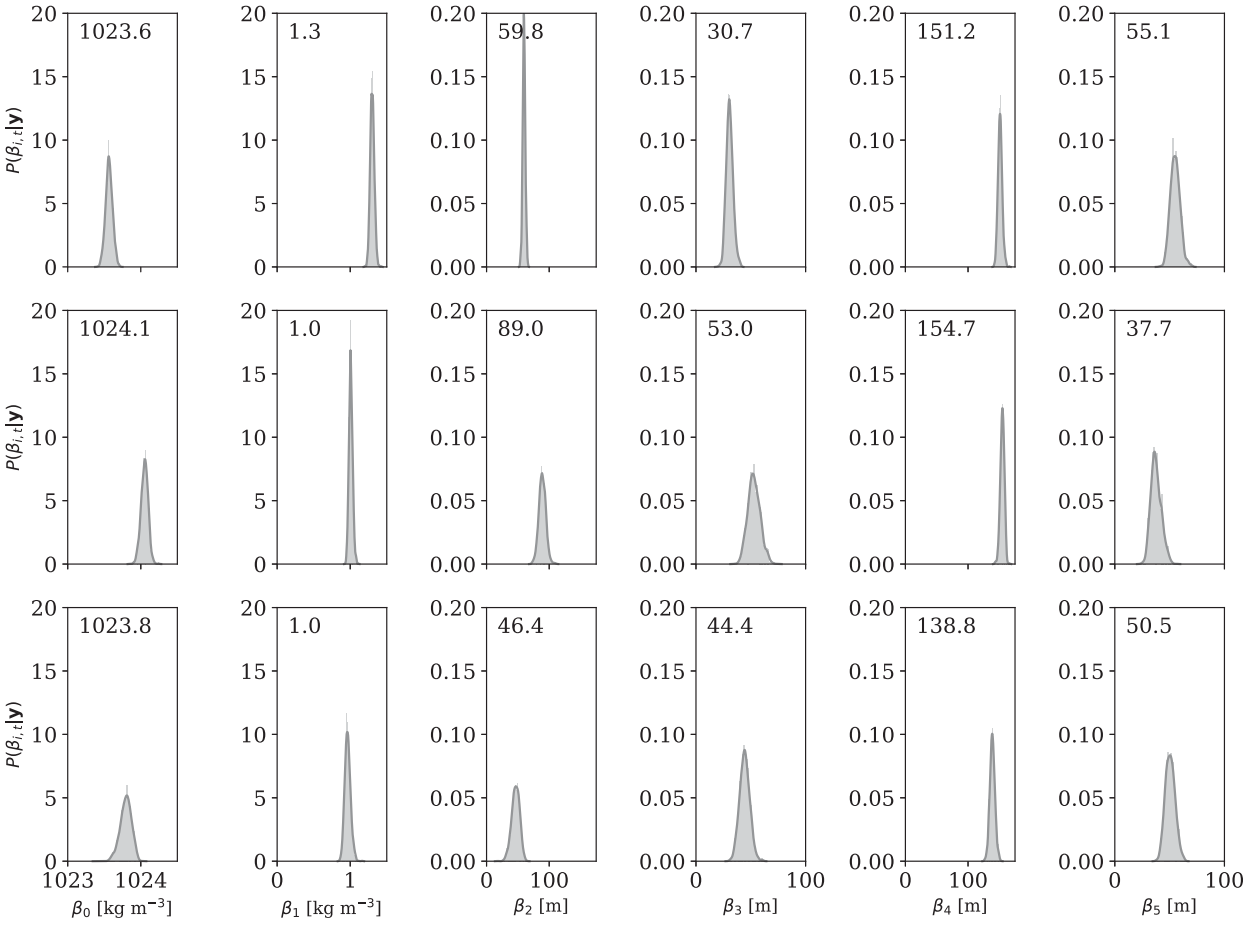

FIG. 7. Uncertainty quantification as represented by the posterior probability distributions of $\boldsymbol{\beta}_t$ for (top) $t = 27$, (6 May 2016), (middle) $t = 469$ (25 Aug 2016), and (bottom) $t = 647$ (8 Oct 2016). The value in each panel indicates the posterior mean.

The $\beta_{1,t}$ had a minimum expected value of 0.90 kg m^{-3}, with a credible interval of (0.81, 1.00), on 2 October 2016 ($t = 621$), and peaked on 3 May ($t = 10$) at 1.30 kg m^{-3} [credible interval of (1.26, 1.36)]. The primary pycnocline $\beta_{2,t}$ was deepest on 16 August 2016 ($t = 308$) with an expected value 129.7 m [credible interval of (121.0, 140.3)]. The credible intervals of $\beta_{0,t}$, $\beta_{1,t}$, and $\beta_{2,t}$ suggest increasing uncertainty of the mean density, density difference, and depth of the upper pycnocline from September to November 2016, before the mooring was serviced. There was considerably greater knockdown during this period resulting in the greater uncertainty range of these density model parameters that were dependent on the upper-water-column measurements.

The expected values of $\beta_{3,t}$ varied from 39.9 m [credible interval of (31.6, 49.8)] in winter to 53.7 m [credible interval of (45.9, 61.9)] in the *austral* summer indicating a wider upper pycnocline during the warmer months. Expected values for the secondary pycnocline depth $\beta_{4,t}$ was greatest in the *austral* winter (169 m) and decreased substantially to 115 m during the transitional season around October. There was no clear seasonal pattern in the width of the secondary pycnocline; the expected $\beta_{5,t}$ ranged from 35 m with a credible interval of (28, 41) to 68 m with a credible interval of (55, 80).

b. *Squared buoyancy frequency and vertical mode structure*

Posterior summaries for $N_t^2(z)$ and $\phi_t(z)$ for the three examples shown in Fig. 6 are shown in Fig. 9. The double pycnocline period ($t = 27$) resulted in two maxima of the expected value of $N_t^2(z)$ (Fig. 9a). Depths of these maxima coincided with expected values for $\beta_{2,t}$ and $\beta_{4,t}$. In all three example time periods, depths where the vertical density gradient was largest [i.e., the maxima of $N_t^2(z)$] also exhibited wider credible intervals at depths of 60, 151, and 50 m in Figs. 9a–c, respectively.

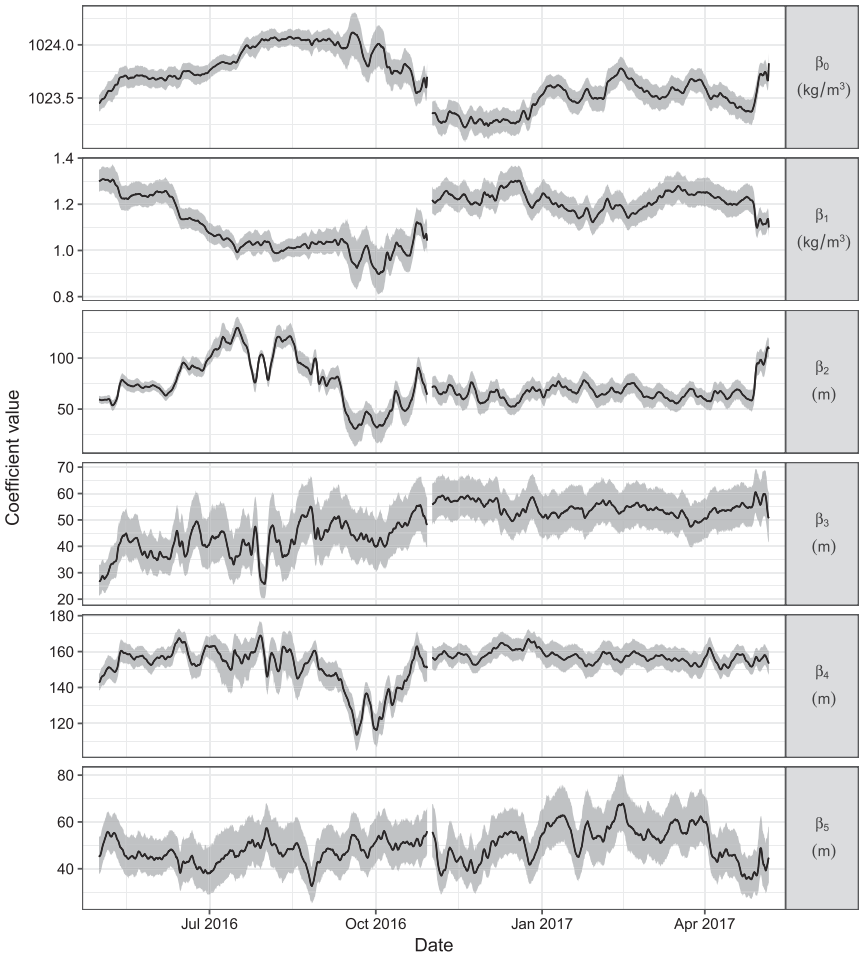

FIG. 8. Posterior distribution for $\boldsymbol{\beta}_t$ for the 12 months of data. For each t the posterior distribution is summarized by the upper and lower bounds of the 0.95 credible interval (gray shading) and the posterior mean (black line).

The credible intervals allow us to make direct probabilistic statements about the properties of $N(z)^2$ and $\phi(z)$. For example, when $t = 27$, the credible intervals only include $N_t^2(z) = 4.0 \times 10^{-4}$ for depths between 48 and 75 m. Outside of these depths, the probability that $N_t^2(z) > 4.0 \times 10^{-4}$ is small. For the corresponding time period, the maximum expected value of $\phi_t(z)$ occurs at a depth of 112 m. However, the credible intervals for $\phi_t(z)$ indicate it is also reasonable that the depth of the maximum lies between 97 and 127 m. Similar probabilistic statements can be made from the profiles during other time periods.

c. *Nonlinear internal wave quantities*

Over the entire observational period, the posterior distributions of phase speed c_t revealed credible interval widths that ranged from 2.5% to 4% of its expected values (Fig. 10a). Credible interval widths for the steepening parameter α_t ranged from 15% to 40% of its expected values (Fig. 10b). As Eq. (5) demonstrates, the steepening length L_s is dependent on both c and α, and note that the width of the credible intervals for the steepening length scale $L_{s,t}$ in Fig. 10c indicate uncertainty levels that varied *by several orders of magnitude*. Given the relatively small credible intervals of the background density profiles (see, e.g., Fig. 6), these results demonstrate large uncertainties in these ocean dynamic quantities.

The credible intervals for α_t provide interesting inference about the nonlinear internal wave characteristics at certain times (Fig. 10b). On occasion (e.g., during October 2016), $p(\alpha_t < 0|\mathbf{y}) \approx 1$, indicating persistent high probability that the environment was conducive to nonlinear internal waves of depression. Alternatively (e.g., during July 2016 and toward the end of the observational period), $p(\alpha_t > 0|\mathbf{y}) \approx 1$, indicating persistent high probability that the environment

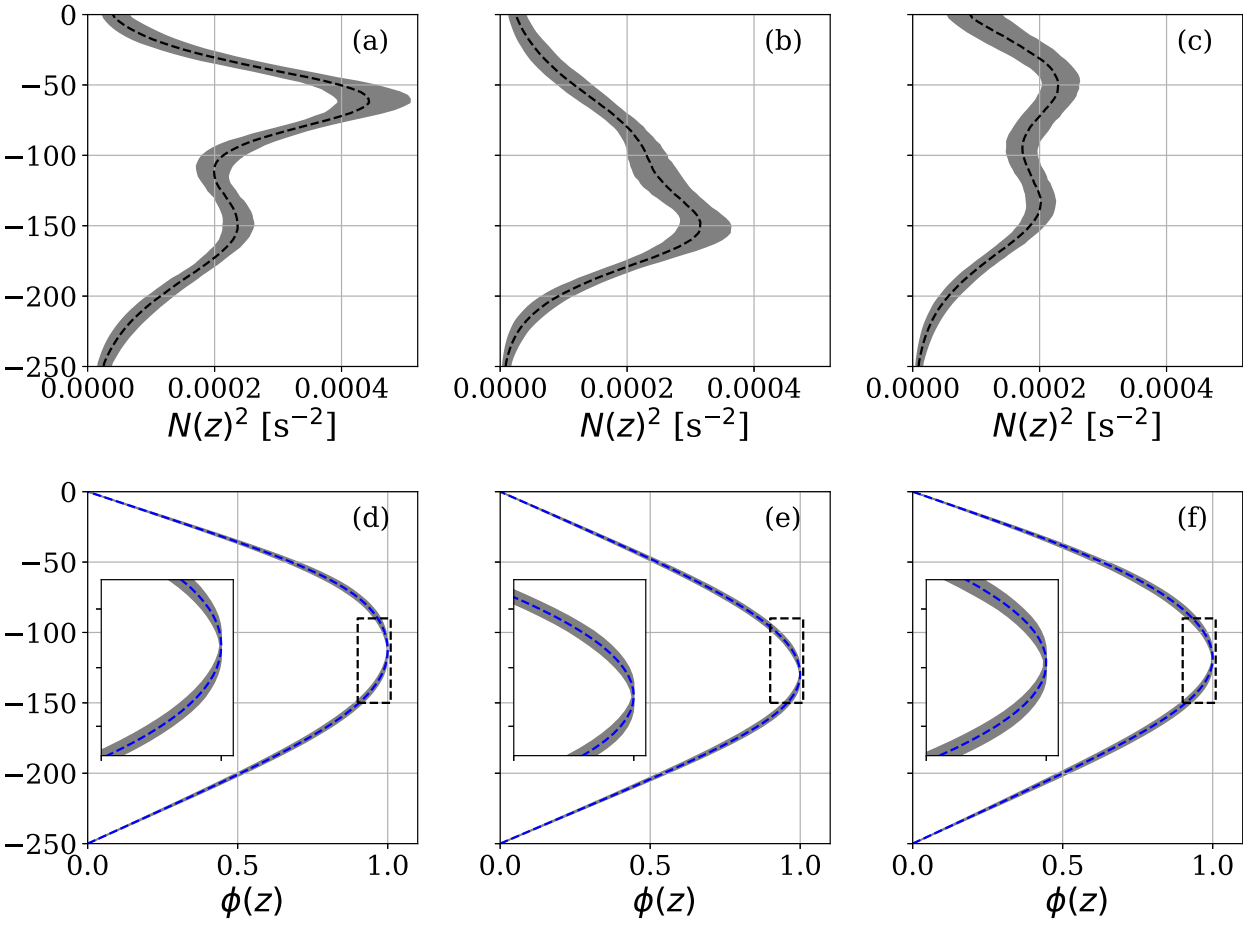

FIG. 9. Posterior mean and 0.95 credible regions for the (a)–(c) squared buoyancy frequency $N(z)^2$ and (d)–(f) mode-1 vertical structure functions $\phi(z)$. Columns correspond to the dates in Fig. 6.

was conducive to nonlinear internal waves of elevation. When the credible intervals contained zero (e.g., 1 December 2016), it was uncertain whether waves of depression or elevation would evolve. When credible intervals contain zero and were narrow (e.g., July 2016), it was most likely that the environment was thus not conducive to any nonlinear internal wave formation.

The steepening length L_s (Rayson et al. 2019) indicates nonlinear internal wave occurrence at a given location in the following manner. If δ is a distance from the internal wave generation site, then $p(L_s < \delta|\mathbf{y}) \approx 1$ means nonlinear internal wave steepening will almost surely occur at the site and solitary-like waves will likely be present. Alternatively, $p(L_s < \delta|\mathbf{y}) = 0$ means nonlinear internal wave steepening will almost surely not occur. At the mooring site, $\delta \approx 100$ km, so we can make statements like $p(L_s < \delta|\mathbf{y}) = 0$ during mid-June 2016 and $p(L_s < \delta|\mathbf{y}) \approx 1$ during October 2016. During certain periods, $0 < p(L_s < \delta|\mathbf{y}) < 1$ (e.g., late March 2017), meaning there is some probability that nonlinear internal waves will form given the credible interval width of L_s.

d. *Isopycnal height*

Figure 11 shows the expected values and credible intervals for isopycnal heights as the density profile evolves over time. The credible interval widths for isopycnal heights ranged from 5 to 10 m at the approximate location of the main pycnocline (i.e., between 50 and 200 m). Uncertainty in isopycnal heights increased near the seabed and surface, where credible interval widths extended to approximately 20–40 m. These are usually regions of weaker vertical density gradients that result in larger isopycnal height uncertainty. Note that contouring via linear interpolation, as is typical in the literature and shown in Fig. 2, does not provide this additional information about the heights of individual isopycnals.

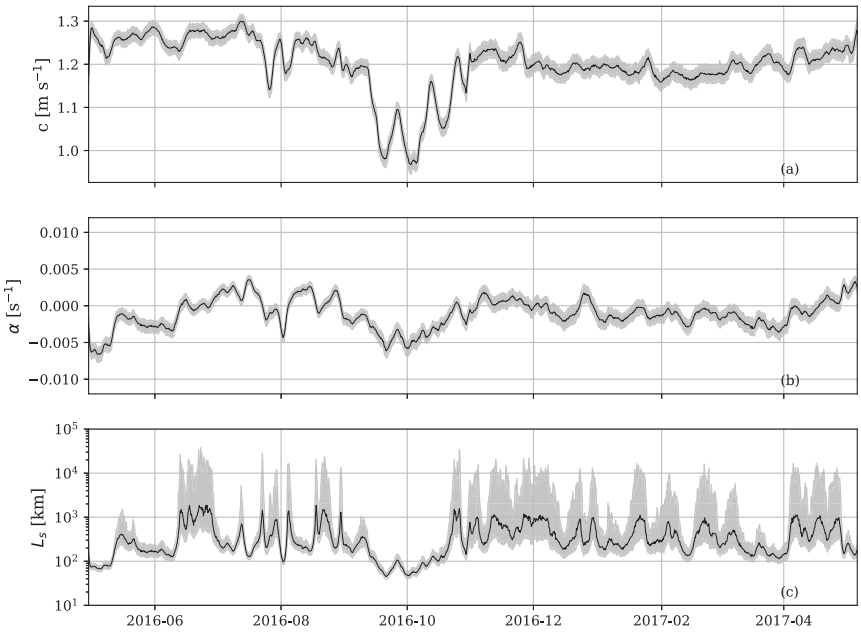

FIG. 10. The 12-month time series of posterior median (black) and 0.95 credible intervals (gray shading) for (a) the linear phase speed c [Eq. (2)], (b) the nonlinear parameter α [Eq. (4)], and (c) the steepening length scale L_s [Eq. (5)]. An initial amplitude $a_0 = 20$ m was used to estimate L_s.

7. Summary and conclusions

We have outlined a methodology for constructing a Bayesian hierarchical model to estimate continuous vertical density profiles using discrete measurements. An analytical function (DHT) was used to objectively characterize the vertical density structure, whereas the BHM with the Hamiltonian Monte Carlo sampling technique infers the posterior distribution of the DHT parameters. The first level of the hierarchy models time-specific background density using a DHT function [Eq. (11)], thereby retaining local temporal characteristics of continuous vertical depth profiles. The second level permits the DHT parameters to vary temporally in order to assess their annual variability over the observational period. When available, scientific knowledge about the behavior of the density characteristics informed the third and final level prior distributions; otherwise, they were allocated weakly informative priors. It was shown that the DHT function is a flexible model for background density, performs well across different stratification conditions, and yields probabilistic statements about vertical density profiles and their characteristics.

For the interested user, variations to the BHM are straightforward to implement in the Stan software provided. For example, rather than the DHT function in Eq. (12), analytical functions such as exponential stratification models or those described in González-Pola et al. (2007) and Zhou et al. (2017) could be used, as could nonparametric mean functions such as smoothing splines or Gaussian processes. Similarly, various prior specifications, including nonparametric priors, over various model parameters may be substituted for the ones mentioned in this article. Furthermore, the DHT can be adapted to model temperature structure, instead of density, as mentioned in section 3.

This article concentrates on uncertainty as quantified by probability. The BHM provided posterior probabilistic inference for upper-ocean density characteristics, as given by the DHT parameters, at both at a given point in time and their evolution over a year. A novel way of assessing and visualizing the uncertainty in isopycnal heights was shown that provides users with information beyond what is typically reported via interpolating contour lines. Background density is also used as an initial condition in the calculation of various ocean scalars, process models, and parameterizations. We have provided a demonstration of how to propagate the uncertainty in initial background density profiles to obtain posterior distributions of various ocean dynamical properties that may serve as instruments for probabilistic inference.

These results also suggest the need for full uncertainty propagation, as opposed to ad hoc sensitivity analyses of

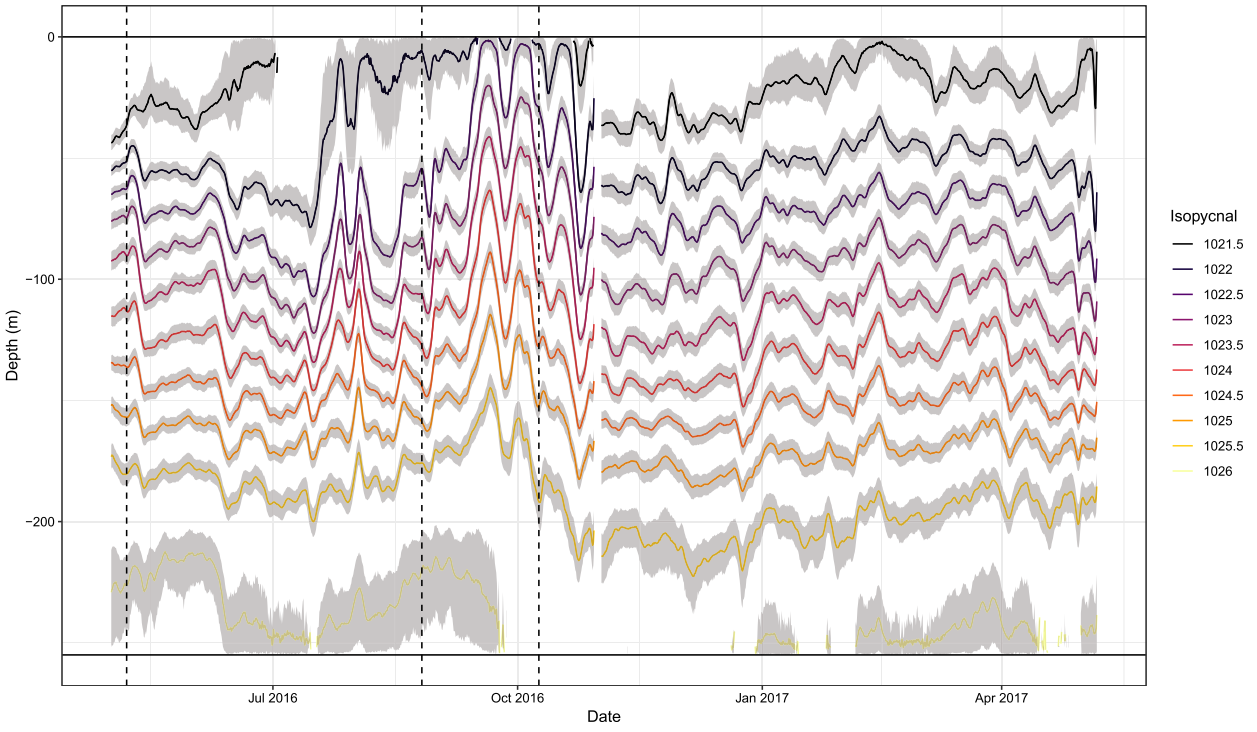

FIG. 11. The temporal evolution of the background density derived from the fitted hierarchical model. The posterior mean (line) and 0.95 credible intervals (gray shading) are displayed for the isopycnal values specified in the legend. The dashed vertical lines correspond to the times at which the density profiles depicted in Fig. 6 occur.

input variables, to fully appreciate the effect that the background density profile uncertainty has on ocean dynamic quantities such as $N(z)^2$, c, and L_s. For example, at times during the observational period, the nonlinear steepening parameter showed significant uncertainty and the steepening length scale displayed credible interval widths that varied by orders of magnitude. To identify such phenomena a full coverage of the input space, and an associated probability distribution, is required. In this work, they were supplied by the BHM, the posterior distribution of which was validated against real measurements.

Acknowledgments. AM, MR, EC, MH, GI, and NJ were supported by the ARC Industrial Transformation Research Hub for Offshore Floating Facilities, which is funded by the Australian Research Council, Woodside Energy, Shell, Bureau Veritas, and Lloyd's Register (Grant IH140100012). MG was supported by EPSRC grants (EP/R034710/1, EP/R018413/1, EP/R004889/1, EP/P020720/1), an EPSRC Established Career Fellowship (EP/J016934/3), and a Royal Academy of Engineering Research Chair. This work is supported by a grant from the Lloyd's Register Foundation on Data-Centric Engineering. MH was supported by a visiting fellowship from the Alan Turing Institute, London.

REFERENCES

Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, and T. K. Berntsen, 2012: Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, **23**, 253–271, https://doi.org/10.1002/env.2140.

Arthur, R. S., S. K. Venayagamoorthy, and J. R. Koseff, 2017: How we compute N matters to estimates of mixing in stratified flows. *J. Fluid Mech.*, **831**, R2, https://doi.org/10.1017/jfm.2017.679.

Astfalck, L., E. J. Cripps, J. P. Gosling, M. R. Hodkiewicz, and I. A. Milne, 2018: Expert elicitation of directional metocean parameters. *Ocean Eng.*, **161**, 268–276, https://doi.org/10.1016/j.oceaneng.2018.04.047.

Berliner, L. M., C. K. Wikle, and N. Cressie, 2000: Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *J. Climate*, **13**, 3953–3968, https://doi.org/10.1175/1520-0442(2001)013<3953:LLPOPS>2.0.CO;2.

Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory*. John Wiley and Sons, 586 pp.

Betancourt, M., 2017: A conceptual introduction to Hamiltonian Monte Carlo. ArXiv, https://arxiv.org/abs/1701.02434.

——, and M. Girolami, 2015: Hamiltonian Monte Carlo for hierarchical models. *Current Trends in Bayesian Methodology with Applications*, CRC Press, 79–101.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng, 2011: *Handbook of Markov Chain Monte Carlo.* CRC Press, 619 pp.

Carpenter, B., and Coauthors, 2017: Stan: A probabilistic programming language. *J. Stat. Software*, **76**, 1–32, https://doi.org/10.18637/jss.v076.i01.

Chelton, D. B., R. A. Deszoeke, M. G. Schlax, K. El Naggar, and N. Siwertz, 1998: Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, **28**, 433–460, https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2.

Chu, P. C., C. R. Fralick, S. D. Haeger, and M. J. Carron, 1997: A parametric model for the Yellow Sea thermal variability. *J. Geophys. Res.*, **102**, 10 499–10 507, https://doi.org/10.1029/97JC00444.

Feistel, R., 2008: A Gibbs function for seawater thermodynamics for −6 to 80°C and salinity up to 120 g kg^{-1}. *Deep-Sea Res. I*, **55**, 1639–1671, https://doi.org/10.1016/j.dsr.2008.07.004.

Fiedler, P. C., 2010: Comparison of objective descriptions of the thermocline. *Limnol. Oceanogr. Methods*, **8**, 313–325, https://doi.org/10.4319/lom.2010.8.313.

Furrer, R., S. R. Sain, D. Nychka, and G. A. Meehl, 2007: Multivariate Bayesian analysis of atmosphere–ocean general circulation models. *Environ. Ecol. Stat.*, **14**, 249–266, https://doi.org/10.1007/s10651-007-0018-z.

Gelfand, A. E., and A. F. M. Smith, 1990: Sampling-based approaches to calculating marginal densities. *J. Amer. Stat. Assoc.*, **85**, 398–409, https://doi.org/10.1080/01621459.1990.10476213.

Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.*, **1**, 515–534, https://doi.org/10.1214/06-BA117A.

——, and J. Hill, 2007: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 625 pp.

——, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, 2013: *Bayesian Data Analysis*. 3rd ed. CRC Press, 675 pp.

——, D. Lee, and J. Guo, 2015: Stan: A probabilistic programming language for Bayesian inference and optimization. *J. Educ. Behav. Stat.*, **40**, 530–543, https://doi.org/10.3102/1076998615606113.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics*. Academic Press, 662 pp.

Girolami, M., and B. Calderhead, 2011: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. Roy. Stat. Soc.*, **73B**, 123–214, https://doi.org/10.1111/j.1467-9868.2010.00765.x.

González-Pola, C., J. M. Fernández-Díaz, and A. Lavín, 2007: Vertical structure of the upper ocean from profiles fitted to physically consistent functional forms. *Deep-Sea Res. I*, **54**, 1985–2004, https://doi.org/10.1016/j.dsr.2007.08.007.

Hastings, W. K., 1970: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109, https://doi.org/10.2307/2334940.

Hennig, P., M. A. Osborne, and M. Girolami, 2015: Probabilistic numerics and uncertainty in computations. *Proc. Roy. Soc.*, **471A**, 20150142, https://doi.org/10.1098/rspa.2015.0142.

Hoffman, M. D., and A. Gelman, 2014: The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, **15**, 1593–1623, http://jmlr.org/papers/v15/hoffman14a.html.

Holloway, P. E., E. Pelinovsky, and T. Talipova, 1999: A generalized Korteweg-de Vries model of internal tide transformation in the coastal zone. *J. Geophys. Res.*, **104**, 18 333–18 350, https://doi.org/10.1029/1999JC900144.

Jaynes, E. T., and O. Kempthorne, 1976: Confidence intervals vs Bayesian intervals. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Springer, 175–257.

Jeffreys, H., 1998: *The Theory of Probability*. Oxford University Press, 470 pp.

Kang, D., and O. Fringer, 2010: On the calculation of available potential energy in internal wave fields. *J. Phys. Oceanogr.*, **40**, 2539–2545, https://doi.org/10.1175/2010JPO4497.1.

King, B., M. Stone, H. P. Zhang, T. Gerkema, M. Marder, R. B. Scott, and H. L. Swinney, 2012: Buoyancy frequency profiles and internal semidiurnal tide turning depths in the oceans. *J. Geophys. Res.*, **117**, C04008, https://doi.org/10.1029/2011JC007681.

Lamb, K. G., and L. Yan, 1996: The evolution of internal wave undular bores: Comparisons of a fully nonlinear numerical model with weakly nonlinear theory. *J. Phys. Oceanogr.*, **26**, 2712–2734, https://doi.org/10.1175/1520-0485(1996)026<2712:TEOIWU>2.0.CO;2.

Lee, F. M., M. R. Allshouse, H. L. Swinney, and P. J. Morrison, 2018: Internal wave energy flux from density perturbations in nonlinear stratifications. *J. Fluid Mech.*, **856**, 898–920, https://doi.org/10.1017/jfm.2018.699.

Liu, A., and D. Benney, 1981: The evolution of nonlinear wave trains in stratified shear flows. *Stud. Appl. Math.*, **64**, 247–269, https://doi.org/10.1002/sapm1981643247.

Llewellyn Smith, S. G., and W. R. Young, 2003: Tidal conversion at a very steep ridge. *J. Fluid Mech.*, **495**, 175–191, https://doi.org/10.1017/S0022112003006098.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, 1953: Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092, https://doi.org/10.1063/1.1699114.

Milliff, R. F., A. Bonazzi, C. K. Wikle, N. Pinardi, and L. M. Berliner, 2011: Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model. *Quart. J. Roy. Meteor. Soc.*, **137**, 858–878, https://doi.org/10.1002/qj.767.

Monnahan, C. C., J. T. Thorson, and T. A. Branch, 2017: Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods Ecol. Evol.*, **8**, 339–348, https://doi.org/10.1111/2041-210X.12681.

Nash, J., M. Alford, and E. Kunze, 2005: Estimating internal wave energy fluxes in the ocean. *J. Atmos. Oceanic Technol.*, **22**, 1551–1571, https://doi.org/10.1175/JTECH1784.1.

Neal, R., 2011: MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, S. Brooks et al., Eds., CRC Press, 113–162.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow, 2006: *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley and Sons, 338 pp.

Phillips, O. M., 1977: *The Dynamics of the Upper Ocean*. Cambridge University Press, 336 pp.

Polzin, K. L., A. C. Naveira Garabato, T. N. Huussen, B. M. Sloyan, and S. Waterman, 2014: Finescale parameterizations of turbulent dissipation. *J. Geophys. Res. Oceans*, **119**, 1383–1419, https://doi.org/10.1002/2013JC008979.

Rayson, M. D., N. L. Jones, and G. N. Ivey, 2019: Observations of large-amplitude mode-2 nonlinear internal waves on the Australian North West shelf. *J. Phys. Oceanogr.*, **49**, 309–328, https://doi.org/10.1175/JPO-D-18-0097.1.

R Core Team, 2018: R: A language and environment for statistical computing. R Foundation for Statistical Computing, https://www.R-project.org/.

Robert, C., and G. Casella, 2013: *Monte Carlo Statistical Methods*. Springer Science and Business Media, 649 pp.

Singpurwalla, N. D., 2006: *Reliability and Risk: A Bayesian Perspective*. John Wiley and Sons, 396 pp.

Sprintall, J., and D. Roemmich, 1999: Characterizing the structure of the surface layer in the Pacific Ocean. *J. Geophys. Res.*, **104**, 23 297–23 311, https://doi.org/10.1029/1999JC900179.

——, and M. F. Cronin, 2010: Upper ocean vertical structure. *Encyclopedia of Ocean Sciences*, Academic Press, 217–224, https://doi.org/10.1016/B978-012374473-9.00627-5.

Stan Development Team, 2018: RStan: The R interface to Stan, version 2.17.3. R package, http://mc-stan.org/.

Stastna, M., and K. G. Lamb, 2002: Large fully nonlinear internal solitary waves: The effect of background current. *Phys. Fluids*, **14**, 2987–2999, https://doi.org/10.1063/1.1496510.

Wikle, C. K., R. F. Milliff, R. Herbei, and W. B. Leeds, 2013: Modern statistical methods in oceanography: A hierarchical perspective. *Stat. Sci.*, **28**, 466–486, https://doi.org/10.1214/13-STS436.

——, A. Zammit-Mangion, and N. Cressie, 2019: *Spatio-Temporal Statistics with R.* CRC Press, 380 pp.

Zhou, C., X. Ding, J. Zhang, J. Yang, and Q. Ma, 2017: An objective algorithm for reconstructing the three-dimensional ocean temperature field based on Argo profiles and SST data. *Ocean Dyn.*, **67**, 1523–1533, https://doi.org/10.1007/s10236-017-1104-x.